Kineograph: taking the pulse of a fast-changing and connected world

R. Cheng, J. Hong, A. Kyrola, Y. Miao, X. Weng, M. Wu, F. Yang, L. Zhou, F. Zhao, E. Chen

Motivation

- Social Networks, like Twitter or Facebook continuously generate huge quantities of data
- For the analysis to have value the data needs to be processed in real-time
- Existing frameworks, such as Hadoop or Graphlab are unable to provide the timeliness guarantees

Key Components

- Distributed in-memory graph storage system
- Graph engine that supports incremental graph mining
- Snapshot mechanism that produces reliable and consistent updates periodically
- Fault tolerance mechanisms

Kineograph

- Ingest nodes take raw data as input and transform it into a sequenced transaction
- They are also responsible for transmitting the transactions to the graph nodes
- Graph nodes store data and perform computations on it

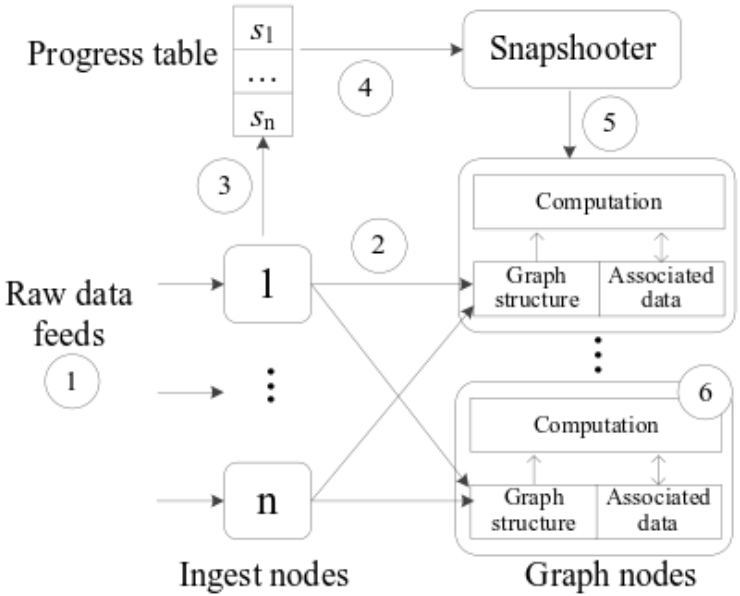
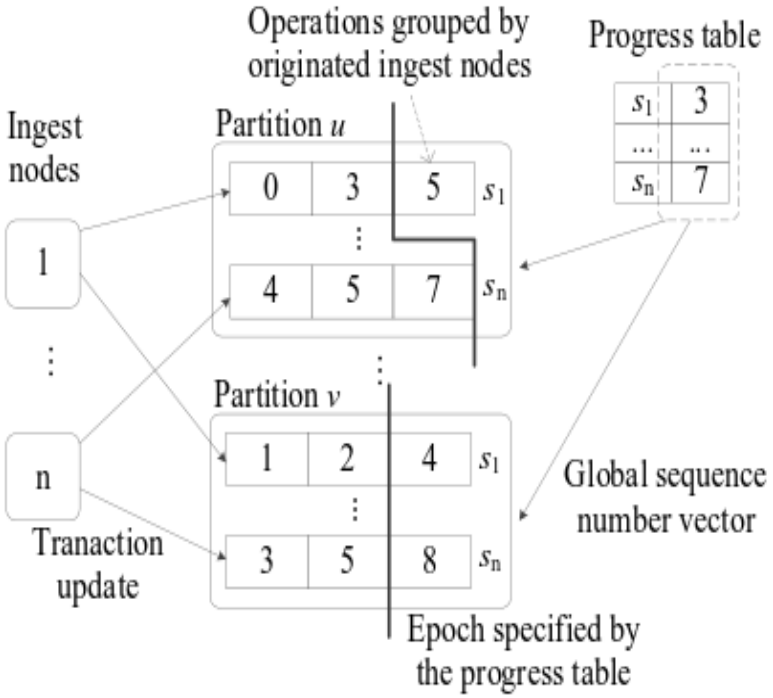


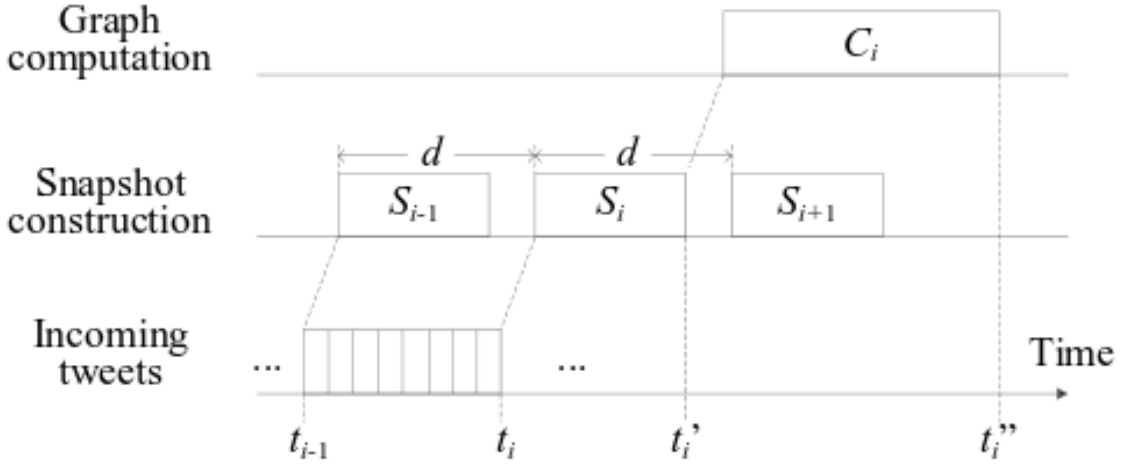
Figure 1. System overview.

Snapshotting

- Kineograph batches operations into small windows, to ensure good timeliness of results

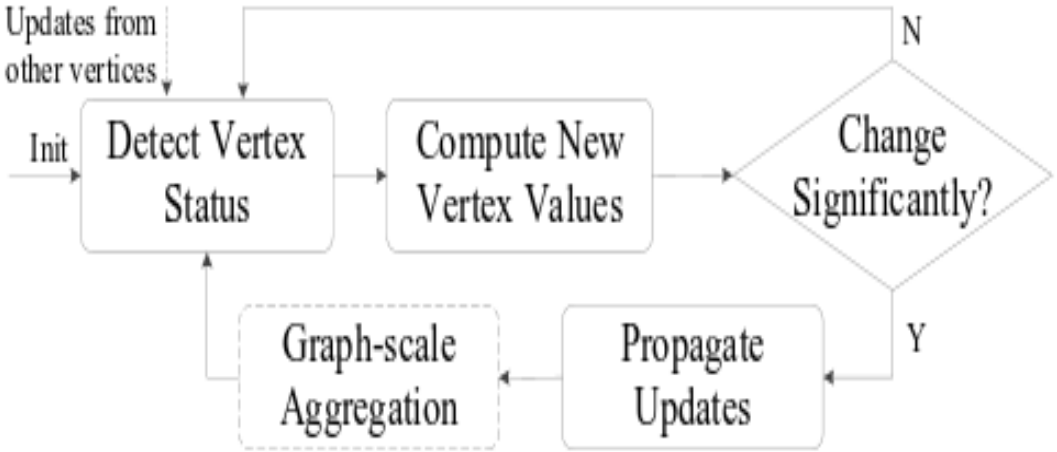


Overview of the computation



Graph Engine

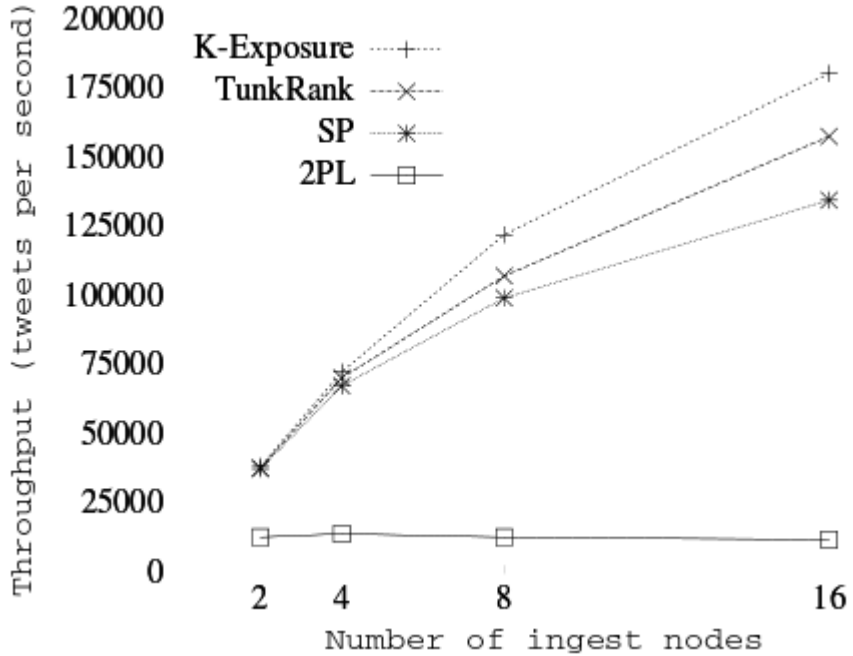
- Familiar vertex centric approach to computation
- Supports both the push (Pregel) and pull (Graphlab) model
- Has support for dynamic computation
- Scheduler does not guarantee sequential consistency (but no write races can occur)



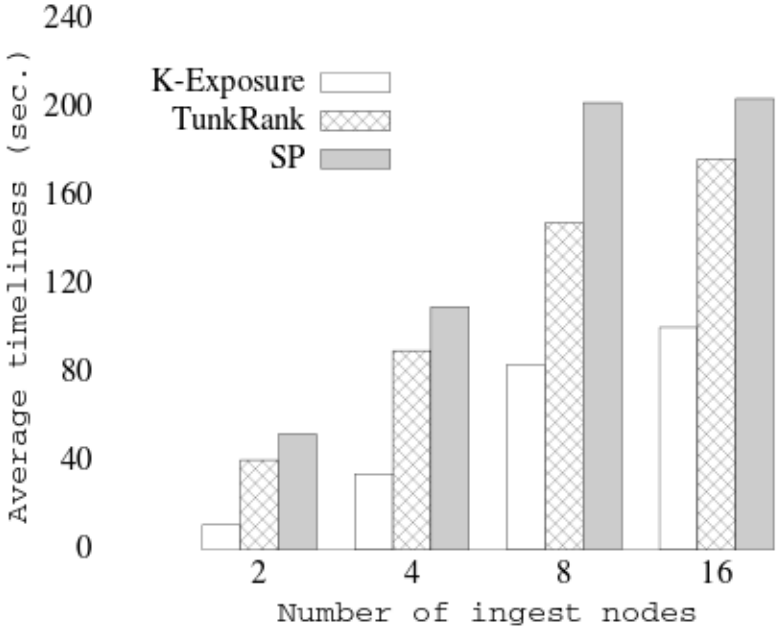
Computation overview

Evaluation: Throughput

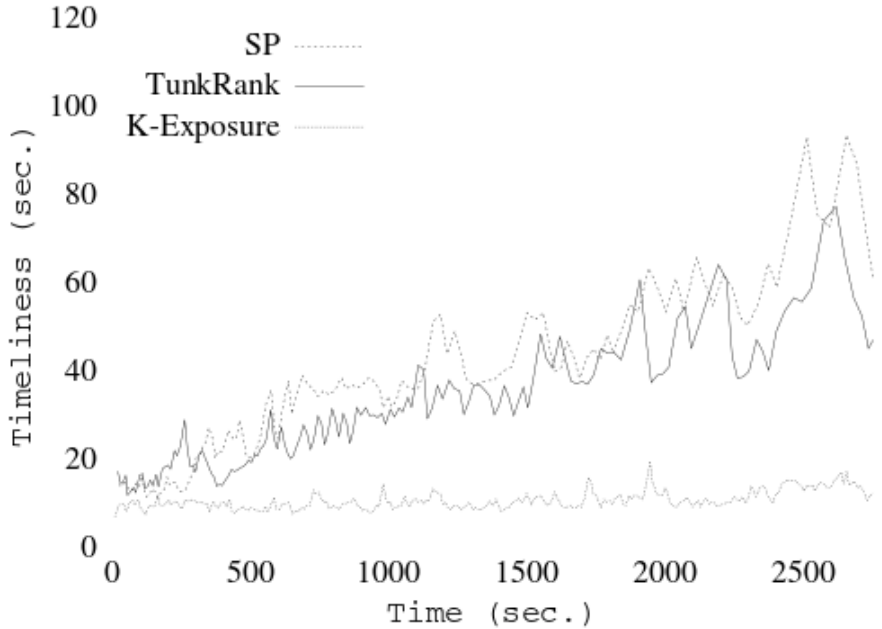
How the throughput of the system varies with the number of ingest nodes



Evaluation: Timeliness



How timeliness varies with the rate of incoming data



How timeliness varies as data is fed into the system

Conclusion

- Graph processing framework, designed for real-time streaming data
- Supports dynamic, incremental graph computation.
- Evaluation leaves a little to be desired, not clear if this framework can run for extended periods of time.

Questions